Credit Scoring in the Context of the European Integration: Assessing the Performance of the Generic Models

by G.Andreeva[1], J.Ansell[2] and J.N.Crook[3]

Management School and Economics, The University of Edinburgh, 50 George Square, Edinburgh, EH8 9JY. Fax +44(0)131-668-3053

Abstract

The on-going European political and economic integration expands the scope of the target population of credit applicants from a single country to the European Union. Traditionally, accept/reject decisions in consumer credit are based on credit scoring models empirically derived for a specific country. The paper addresses the possibility of applying a single generic model to score the population of several European countries and the impact of segmentation (i.e. building individual models for different nations) on predictive accuracy. In contrast to previous studies, the logistic regression generic model built for three European countries shows the adequate performance, comparable to that of the national models. In addition, several practical problems associated with the use of a single model are considered: different strategies for selecting the acceptance levels; the necessity for further harmonisation of data collection practices across the countries; implications for consumers.

Keywords: banking, credit scoring, decision analysis

1. Introduction

With the adoption of the EURO by 12 countries exchange rate risk associated with repayments on credit products was removed and so the cost of using a credit product, for example a card, issued in an institution in a different country to facilitate transactions (possibly in a third member country) was reduced. According to Eurostat (2003), the estimated population within the existing EURO-zone will be nearly 310 million people in 2010. This offers potentially an extremely attractive possibility of market expansion for creditors, provided the default risk of not only the residents of their own country, but also applicants from the neighbouring Member States can be assessed.

[1] Corresponding author: 12/5 Blacket Avenue, Edinburgh, EH9 1RS, UK. Tel.+44(0)131-651-1235. E-mail: G.Andreeva@ed.ac.uk

[2] Tel.+44(0)131-650-3806. E-mail: Jake.Ansell@ed.ac.uk

[3] Tel.+44(0)131-650-3802. E-mail: Jonathan.Crook@ed.ac.uk

Risk assessment techniques in consumer credit, known as credit scoring, are based on historical performance of past borrowers. They relate the default probability to characteristics that can be observed at the point of application (age, marital status, residential status, etc) or can be derived from the lender's records or from credit bureau (number of repayment delinquencies in the past, etc). The accept/reject decision is achieved by setting a cut-off level corresponding to a certain value of the estimated default probability, applicants with default probabilities above this level are not granted credit. For a more detailed overview on credit scoring see Crook (1997); Hand (1998); Thomas et al. (2002).

Most commonly credit scoring models are estimated and applied to one population (e.g. residents of one country that use a particular credit product). An alternative to such customised models is a generic model, which is developed on one or several populations or portfolios and applied to score a geographically or socio-economically different population/ populations (Thomas (2000)). The possibility of constructing a generic model for several different populations - or countries in our case – without significant loss in the quality of prediction follows from the 'curse of insensitivity' (Rapoport (1975), von Winterfeldt and Edwards (1982)), or flat maximum effect (Lovie and Lovie (1986)), which implies that the predictive ability of linear models is insensitive to relatively large variations in regression weights. It follows that different linear models can give the same level of classification accuracy.

The majority of published empirical tests demonstrate the superiority of customised over generic models, see Overstreet et al. (1992), Overstreet and Bradley (1996), Platts and Howe (1997), Staten (1999), Avery et al. (2000), Barron et al. (2000). But on the other hand, Banasik et al. (1996) argue that segmentation is not necessarily going to improve the predictive accuracy, subpopulations need to be 'sufficiently different' and large in size to justify the development of separate scorecards.

In the European context there was only one attempt to construct a Pan-European scorecard - Platts and Howe (1997). The study showed that a European generic model was less predictive compared to customised national models, and less predictive again compared to customised portfolio scorecards.

This paper addresses the same problem in more detail for three different European countries using more recent data. The necessity to revisit this problem stems not only from increasing European cohesion, but also from tightening consumer credit regulations and general anti-discrimination provisions. A possible interpretation of anti-discrimination law may lead to a conclusion that generic model is the only legal model in the EU context, since it does not discriminate between different nationalities (for a more detailed discussion see Andreeva et al. (2004)).

The next section describes the data and methodology used in the analysis. Section 3 comments on the national differences in credit risk patterns observed for the countries under consideration. Section 4 presents the predictive performance of the generic model. The following section investigates the classification accuracy of generic and national models under different cut-off levels. Section 6 comments on differences between applicants accepted by different models. This is followed by the conclusion.

2. Data description and methodology

The data for analysis were provided by a major international credit scoring consultancy and relate to the same retail card issued to applicants from three European countries: Belgium, Germany and the Netherlands. The populations of card applicants differed in size and the period of time for which the performance was recorded and the summary is presented in Table 1.

TABLE 1 HERE

The same application period of 14 months for all three countries was selected to cover the period from November 1998 until December 1999. The performance of the accounts accepted within this period was observed until November 2000. The definition of bad was chosen to be ‘at least 2 months in arrears’ at any time during the observation period. The more traditional definition of bads of being ‘at least 3 months in arrears’ did not give sufficient numbers of bads to allow for effective classification.

All three samples of accepted applications were split randomly into training (70%) and holdout (30%) datasets. The list of characteristics collected in each country was different. However, it was possible to select 16 variables that were collected for all three countries, see Table 2.

TABLE 2 HERE

Characteristics were categorised or coarse-classed on the basis of the weights of evidence (WOE):

$w_{ij} = log\ (g_{ij}B_j/b_{ij}G_j)$,

where g_{ij} (b_{ij})are the corresponding numbers of goods and bads within the attribute i of characteristic j, G_j (B_j) are total numbers of good/bad in the sample.

The following model was fitted to three national training samples :

$$\log[\frac{p_{ij}}{1-p_{ij}}] = \alpha_j + \beta_{j1}x_{ij1} + \beta_{j2}x_{ij2} + ... + \beta_{jk}x_{ijk}$$

where p_i is the probability of being good for individual i, country j, and predictor variable k. Stepwise selection of variables was used. Then the generic model was built on the aggregated training sample, and its predictive performance was compared to that of national models.

The four models described above were developed on a set of variables that was common to all three countries. In addition, three models were developed on a full set of characteristics that was available for each country, and the resulting improvement in predictive performance was measured.

The performance was measured on hold-out samples using area under the receiver operating characteristics curve (AUROC), see Bamber (1975). A ROC-curve is a plot of the proportion of correctly classified good accounts against the proportion of incorrectly classified bad accounts for all levels of the score or predicted default probability. It follows that this measure is independent from the cut-off. The higher the curve, the better is prediction, and consequently the larger is the area under the curve, the better is the model performance. Conceptually the area under the ROC-curve corresponds to the Wilcoxon or Mann-Whitney or U statistic. This statistic estimates the probability, θ, that a ranking of a randomly selected bad account will be less than or equal to a ranking of a randomly selected good account (Hanley and McNeil (1982)).

Whilst AUROC offers a convenient measure of predictive accuracy for all possible cut-offs or acceptance levels, in real life it is necessary to select a certain cut-off, and hence to decide how many applicants will be granted credit. The second measure used for comparison is based on the confusion matrix, which presents the counts of good and bad accounts correctly and incorrectly classified by the model for a chosen cut-off level.

TABLE 3 HERE

Following the approach of Banasik et al. (1996), we will set a cut-off at a fixed level of either b (the number of goods predicted as bads) or $b+d$ (the total number predicted as bads, and therefore, rejected by the model). In both cases, d (the number of correctly classified bads) will provide a measure for comparison of predictive accuracy of models.

3. National differences

The data revealed different patterns for each country.

Figure 1 gives an example of 'Marital Status', the chart shows weights of evidence for each attribute of this characteristic.

FIGURE 1 HERE

Examination of the chart shows that:

- There existed unique attributes that were not present in all countries (e.g. 'Living Together Registered' was specified as a separate category of 'Marital Status' for the Netherlands but did not exist for Belgium or Germany).
- The level of risk for the same attributes differed across the countries.

At the same time it was possible to observe some general patterns that appear to hold irrespective of the country. For example for all three countries married and widowed applicants were the categories with the lowest probability of default, single applicants had the highest probability, and divorced / living together appeared in the middle.

The variables selected by the stepwise routine and their ranking look very different. Table 4 gives the characteristics that were significant for different countries ranked in the order of decreasing significance. The characteristics of the type 'Occupation*Business type' present the variable constructed of two characteristics. This was done to avoid high degrees of collinearity between these characteristics.

TABLE 4 HERE

To summarise, the disparity between the nations was evidenced by differences in

1. predictive performance
2. attributes for each characteristic and in coarse-classed aggregations of the attributes;
3. coarse-classes selected by the stepwise routine;
4. different WOE and β-values for any common attributes or coarse-classes.

4. Generic model

The generic model was fitted on a sample from three national datasets used in the analysis above. First we estimated a model using all of the variables in Table 2 but without any variables to indicate which country the case was a member of. Second we included country indicators as dummy constants. The inclusion of the country identifying dummies slightly improved the predictive ability of the model.

To test the predictive performance, the generic model was applied to the aggregated hold-out sample and also to each of the national segments within the hold-out in order to compare its predictive performance against each of the national models separately. The results, presented in Table 5, indicate that the generic model performs well, showing only minor loss in the quality of prediction as compared to national models.

TABLE 5 HERE

Comparisons of the generic model with national ones would be incomplete without incorporating full information that was available for each country into the analysis. A third type of model was built – national models developed on a full set of characteristics, both application and bureau that were available for each country in turn. These models contained different variables and coarsely classed characteristics between the countries.

The performance of generic, national and full-information models is summarised in Table 5 and Figure 2. For Belgium 8 additional variables entered the model, for the Netherlands – 13, among them 6 credit bureau variables, for Germany – 23, among them 8 bureau variables.

FIGURE 2 HERE

In contrast to previous findings, the generic model built for different populations of three European countries showed the adequate performance comparable to that of the national models using the same variables. But incorporating nationally specific additional information into the models increases their predictive ability further. This was especially evident for the Netherlands and Germany, where quite a lot of new variables, including several bureau characteristics, entered into the model. So the major difference between generic and customised models stems from the scope of information available for analysis rather than from different grouping and weighting of similar characteristics. Thus generic scoring is potentially a viable option provided that the data collection is harmonised between the countries.

We can think of the following reasons for the similarity in predictive performance between the generic and national models. The countries used in the analysis are geographically and economically similar. The results may be different if geographically and economically dissimilar countries are included, for example Ireland and Italy. Nevertheless it is possible to argue that generic scoring is possible for some parts of Europe.

Although there were differences in distributions of good/bad classes across the countries, there were general patterns that could be observed, e.g. married applicants were better credit risks than single ones, homeowners were better then those in rented accommodation, older applicants were better than the younger ones. There were no strong interactions between the predictor variables and country indicators, and the observed variation across the countries was compensated by the flat maximum effect. In addition, the generic model was developed on a heterogeneous sample, in which all three subpopulations were adequately represented.

5. Predictive performance under different cut-off levels

In order to arrive to an accept/reject decision, it is necessary to decide on a threshold that will determine the acceptance rate. We consider three alternatives, and investigate the predictive accuracy of generic and national models built on the same set of characteristics.

1) Fixing the number of goods predicted as bads
   a) For all scorecards the cut-off is set so that number of goods predicted as bads is 10% of the total number of actual goods. The results are reported in Table 6, columns 5-12.
   b) Whilst the above presented approach offers an independent way of setting cut-offs across generic and national scorecards, 10% error rate may not necessarily by an optimal level for the national scorecards. It may be that for different countries the lower or higher values of error rate may give better numbers of correctly classified bads. To overcome this difficulty the numbers of goods predicted by each national models were fixed to equal that number predicted by generic model with 10% error rate, see Table 6, columns 5-8 and 13-16.
2) Fixing the total number of predicted bads
   a) One way of selecting a cut-off that will be independent across all scorecards is to fix the probability level for all scorecards. It was chosen to set a cut-off at the probability level predicted by the generic model (0.8826) so that the number of predicted bads (or rejects) would be close to 40% of total sample, an approximate average reject rate across three countries. The results are presented in Table 7, columns 2 and 4-7.
   b) A similar approach, but a cut-off level for each scorecard is fixed to give the number of predicted bads that is equal to the number of actual bads in each national segment of the hold-out sample (Table 7, columns 3 and 9-12).
3) Marginal Good:Bad ratios

Normally the cut-offs are chosen by lenders in such a way that the loss from bad accounts do not exceed the profit from good accounts, it can be expressed by a Good:Bad ratio, i.e. the number of good accounts required to offset the loss from one bad account. Since no information on profit/loss was available for analysis, we decided to select marginal ratios so that they would give roughly the same reject rates as were observed in the sample. Figure 3 shows that marginal ratio of 7:1 will give an approximately 37% reject rate, which the lowest rate observed in the sample, and marginal ratio of 9:1 will give an approximately 45% reject rate, the highest observed level. For generic and national scorecards the cut-off is chosen at the probability (score) level where the marginal ratio changes to a lower integer value. Since the ratio does not change linearly with the score, there is a certain degree of arbitrary judgement involved. The results for marginal cut-offs are reported in Table 8.

TABLES 6,7,8 ABOUT HERE

FIGURE 3 HERE

The last row in Table 7 shows the difference between the number of bads correctly identified by national models and the same number for generic model. The generic model demonstrates more powerful discrimination under both 1(a) and 1(b) cut-off levels, i.e. it identifies correctly more bads for the same number of correctly classified actual goods. It is superior again in predicting bads under 2(a) fixed probability cut-off, as shown in column 8 of Table 8, but it rejects more good accounts. If we allow the cut-off to depend on the bad rate in each national sample, the national models correctly classify 62 customers more than the generic model. However, this constitutes only 0.1% from the total hold-out sample.

Marginal ratio 7:1 gives an advantage to the generic model in identifying bads, but national models reject less goods, and with marginal ratio 9:1 national models show overall superiority, but this comes at the expense of 45% reject rate.

In general, the choice between the generic and national models would depend on the values each lender associates with two types of error: an error of rejecting a good account and an error of accepting a bad account. Under a range of different cut-offs the generic model proved to be superior in terms of the second type of error. We would also like to emphasise that generic models are cheaper to develop and maintain, and this should make them even more attractive.

6. Differences in applications accepted by generic or national models

In this section the view of the applicant will be considered, and it will be shown how different models affect chances of being accepted for certain applicant groups. To do so, it is necessary to investigate what characteristics distinguish between the two groups:

1. the applicants predicted ‘Good’ (and therefore accepted) by the generic model and predicted ‘Bad’ (and rejected) by the country model;
2. the applicants predicted ‘Good’ by the country model and ‘Bad’ by the generic model.

For each national segment in the hold-out sample such groups were selected under cut-off 2(b). For Belgium these groups constituted 3.83% each from the corresponding national segment, for the Netherlands 2.75%, and for Germany 2.82%. In other words, the decision whether to grant credit or not would be identical (irrespective of whether the generic or national decision rule is used) for 92.34% of Belgian applicants, 94.50% of the Dutch, and 94.36% of German applicants.

So for each country the two groups for whom the decision would be different were cross-examined by looking at frequency distributions for categorical characteristics, and means and medians for continuous characteristics. Table 9 presents the characteristics that showed the most striking differences for Belgium. For categorical characteristics only some attributes are selected for the purpose of illustration, that is why the percentages within one characteristic do not always sum to 100%. As a benchmark, the frequency distributions, means and medians are also reported for the total national segment.

TABLE 9 HERE

Applicants accepted by the national model, but rejected by the generic model would be slightly younger than those accepted by the generic model, but rejected by the national one: a median age for former would be 25 years versus 29. Those working in service sector or self-employed, and those living with parents would prefer to be scored by a national model, because they will have then a higher chance of acceptance. And the generic model favours people working in industry, with no card insurance, divorced and renting a flat. So for the consumer it does make a difference which model is applied.

7. Conclusions

This paper presented an investigation of risk patterns in three European countries and compared the performance of the generic model to that of models built for each country separately. We find that the generic model showed an adequate predictive performance comparable to that of its national counterparts, both under measures independent of any cut-off level and under different cut-off strategies. Such a performance can be attributed to the relative similarity of the countries used in the analysis, which meant that it was possible to select a set of characteristics that could be harmonised across three countries. This set contained enough information to allow for good classification. There were general patterns that could be observed in distributions of good/bad classes across the countries, and the differences were compensated by the flat maximum effect.

Given the fact that generic models are less costly in terms of development and maintenance, it is possible to conclude that generic scoring is a viable and attractive option for some parts of Europe. Further research is required to establish whether the results of this study can be extended to other European countries.

Nevertheless, whilst the generic model showed an acceptable level of predictive performance, the applicants accepted with the generic decision-rule differed from those accepted with the national-specific rule on a number of characteristics. This highlights a potential problem of legal compliance of credit scoring models. Given a diversity of national laws and a lack of a clear definition what constitutes a 'fair' credit granting decision, see Andreeva et al. (2004), the model perfectly legal in one European country, may appear illegal in the neighbouring country. We would like to stress the importance of one common set of rules in EU in relation of legal compliance of credit scoring models.

The paper demonstrates that additional information increases the predictive ability of models, whilst the difference between generic and national models developed on the same set of common characteristics is marginal. This emphasises the value of information and the need for harmonisation of data across Europe. The generic model presented in this paper does not include credit bureau characteristics, since they are not harmonised at present. One may expect that further European harmonisation will increase the potential scope of application of generic models.

References

Andreeva, G., Ansell, J., Crook, J.N., 2004. The impact of anti-discrimination laws on credit scoring, Journal of Financial Services Marketing, 9 (1) 22-33.

Avery, R.B., Bostic, R.W., Calem, P.S., Canner, G.B., 2000. Credit scoring: statistical issues and evidence from credit-bureau files, Real Estate Economics, 28 (3) 523 - 547.

Bamber, D., 1975. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph, Journal of Mathematical Psychology, 12 387 - 415.

Banasik, J.L., Crook, J.N., Thomas, L.C., 1996. Does scoring a subpopulation make a difference?, The International Review of Retail, Distribution and Consumer Research, 6 (2) 181 - 195.

Barron, J.M., Ellienhausen, G., Staten, M.E., 2000. Monitoring the household sector with aggregate credit bureau data, Business Economics, 35 (1) 63 - 77.

Crook, J.N., 1997. Application credit scoring: An overview, Journal of Financial Services Marketing, 2 (2).

Eurostat, 2003. Eurostat Yearbook 2003, http://epp.eurostat.cec.eu.int/cache/ITY_OFFPUB/FREESELECT2/EN/FREESELECT2-EN.PDF.

Hand, D.J., 1998. Consumer credit and statistics. In: Hand,D.J., Jacka,S.D. (Eds.), Statistics in Finance, Edward Arnold, London, pp. 69-82.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1) 29 - 36.

Lovie, A.D., Lovie, P., 1986. The flat maximum effect and linear scoring models for prediction, Journal of Forecasting, 5 (3) 159 - 168.

Overstreet, G.A.J., Bradley, E.L., Kemp, R.S., 1992. The flat-maximum effect and generic linear scoring models: a test, IMA Journal of Mathematics Applied in Business & Industry, 4 (1) 97 - 109.

Overstreet, G.A.J., Bradley, E.L.J., 1996. Applicability of generic linear scoring models in the U.S. credit-union environment, IMA Journal of Mathematics Applied in Business & Industry, 7 291 - 311.

Platts, G.,Howe, I., 1997. A single European scorecard. Credit Scoring and Credit Control V, Conference Proceedings, Management School, Edinburgh.

Rapoport, A., 1975. Research paradigms for studying dynamic decision behaviour. In: Wendt,D., Vlek,C.A.J. (Eds.), Utility, Probability and Human Decision Making, Reidel, Dordrecht.

Staten, M.E., 1999. Strength in numbers: aggregate credit bureau data benefit portfolio management, The Journal of Lending & Credit Risk Management, 82 (1) 27 - 31.

Thomas, L.C., 2000. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers, International Journal of Forecasting, 16 (2) 149 - 172.

Thomas, L.C., Edelman, D.B., Crook, J.N. 2002. Credit scoring and its applications, Society for Industrial and Applied Mathematics, Philadelphia, PA.

von Winterfeldt, D., Edwards, W., 1982. Costs and payoffs in perceptual research, Psychological Bulletin, 91 609 - 622.

Figure 1. Coarse-classification of 'Marital Status' by country

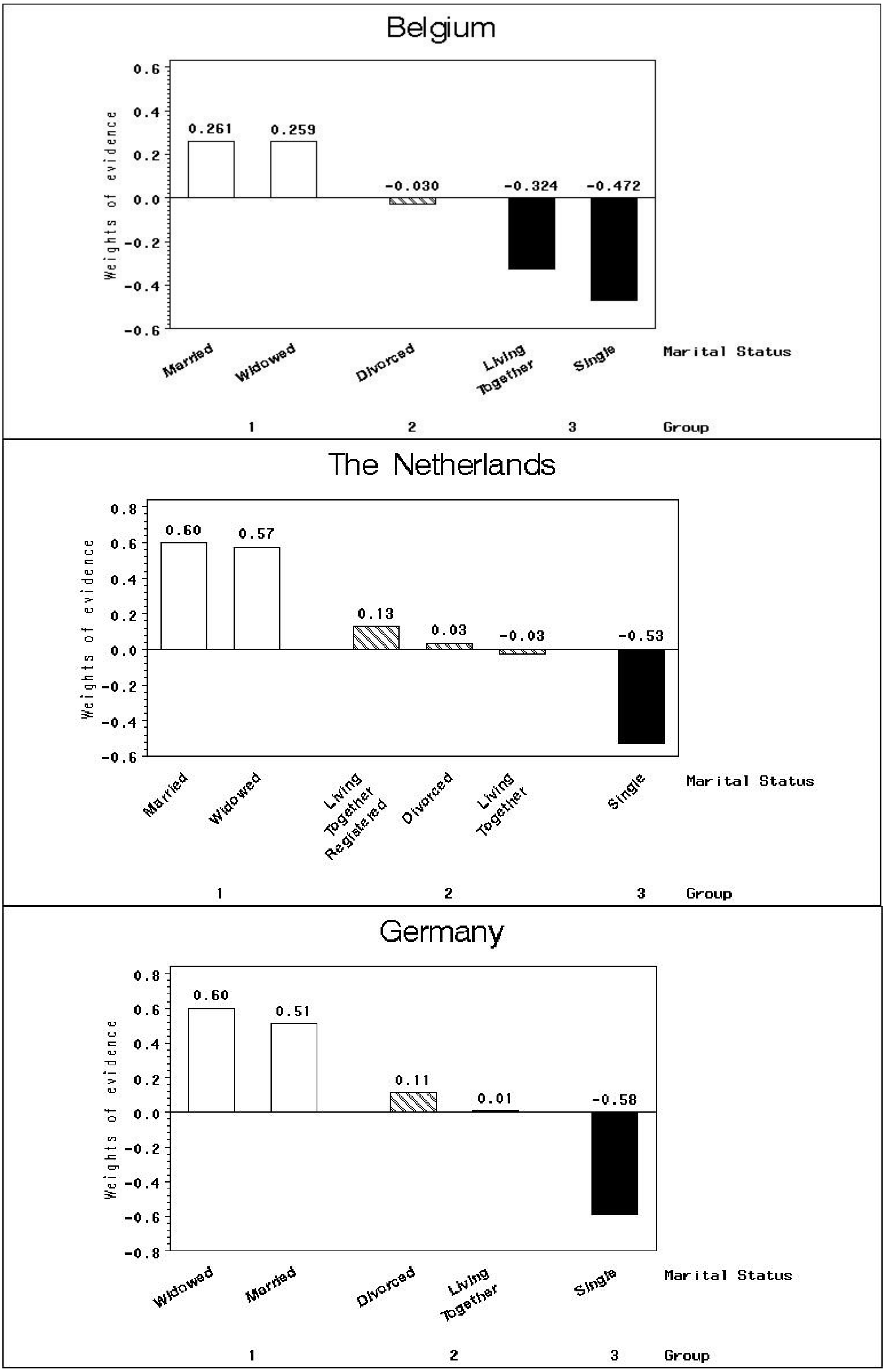

Figure 2. Predictive performance of generic, national and full-information models tested on national hold-out samples.

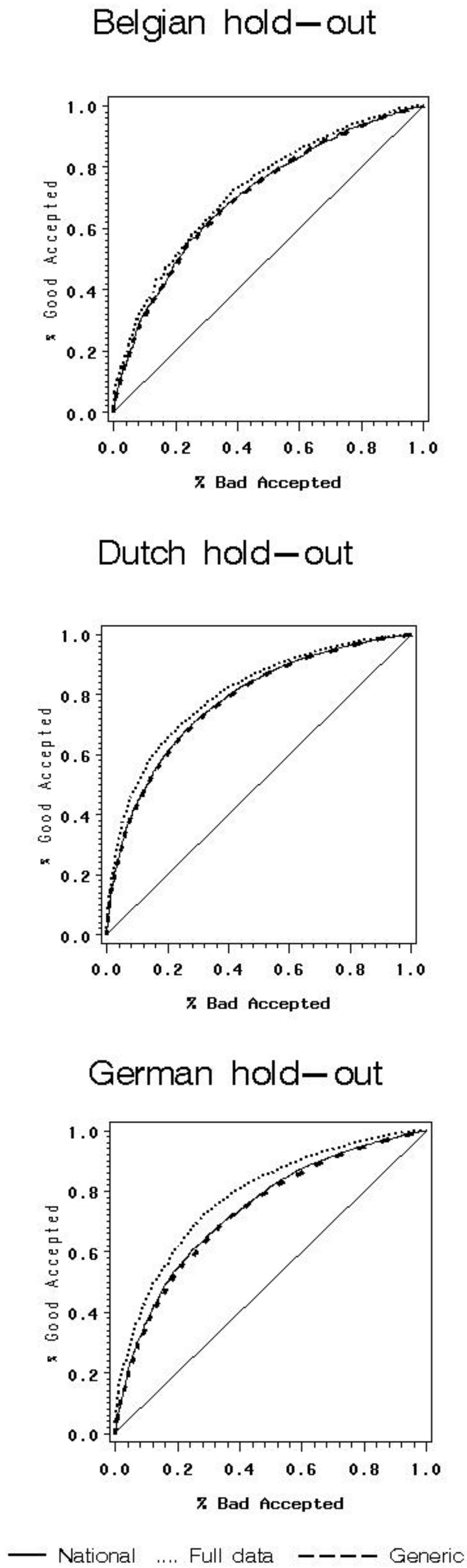

Figure 3. Marginal ratio by score and reject rate

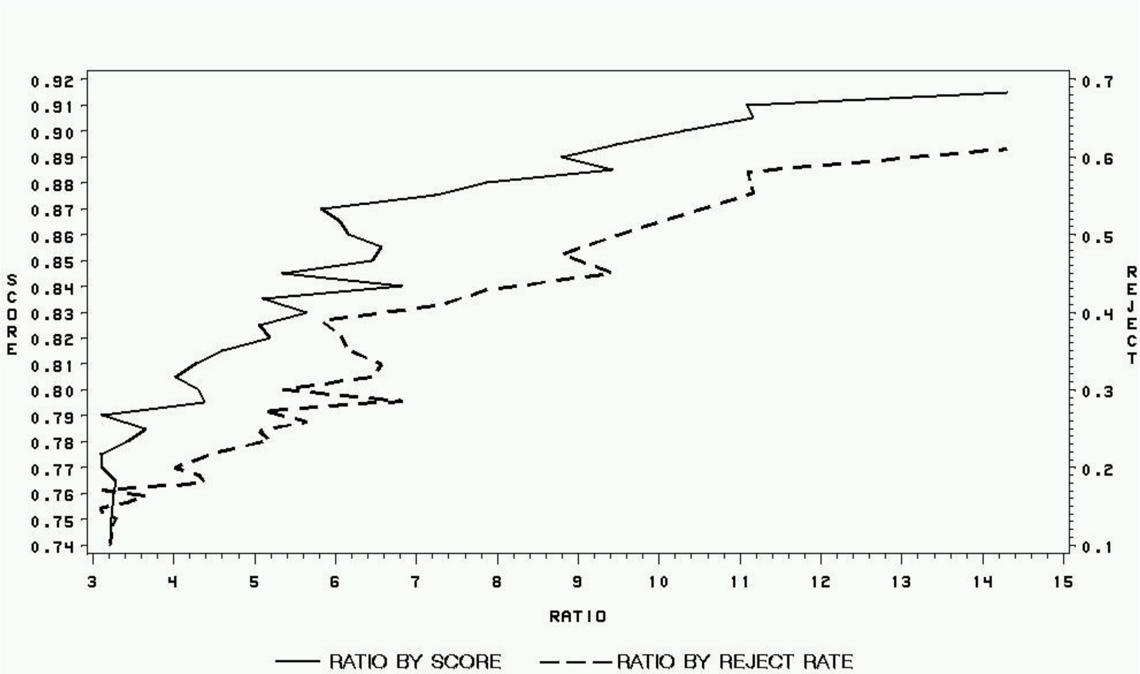

Table 1. Samples used in the analysis.

	Belgium		**The Netherlands**		**Germany**	
	Count	*Percent*	*Count*	*Percent*	*Count*	*Percent*
Bad	3090	6.98%	11213	8.55%	8909	6.50%
Good	23200	52.42%	71024	54.13%	66939	48.81%
Rejected	17966	40.60%	48953	37.32%	61288	44.69%
Total	44256	100.00%	131190	100.00%	137136	100.00%
No of periods (months)	**26** from 01/10/1998 to 30/11/2000		**35** from 01/01/1998 to 30/11/2000		**68** from 01/04/1995 to 30/11/2000	
No of characteristics	**38**		**60**		**75**	

Table 2. List of common characteristics

No	Characteristic	No	Characteristic
1	Home telephone	9	Employer's phone
2	Residential status	10	Card insurance
3	Marital status	11	Credit insurance
4	Occupation (Full-time, part-time, self-employed, etc.)	12	Number of dependants
5	Age	13	Spouse age
6	Time at address since 18 years old	14	Goods code
7	Time in employment	15	Goods price
8	Type of business (Manufacturing, banking, catering, etc.)	16	Payment date

Table 3. Example of confusion matrix

		Predicted Class	
		Good	Bad
Actual Class	Good	**a**	**b**
	Bad	**c**	**d**

Table 4. National models: ranks of characteristics and predictive performance

Rank	Belgium	The Netherlands	Germany
1	Occupation	Business type	Business type
2	Number of dependants	Telephone	Occupation
3	Time at address	Goods code*Payment date	Spouse age
4	Residential status	Marital status	Applicant's age
5	Goods code	Applicant's age	Telephone
6	Payment date	Occupation	Time on job
7	Goods price	Residential status	Goods code
8	Telephone	Number of dependants	Number of dependants
9	Time on job	Time at address	Time at address
10	Spouse age	Credit insurance	Card insurance
11	Business type	Goods price	Residential status
12	Applicant's age	Time on job	Payment date
13	Credit insurance	Spouse age	Marital status
14	Employer's phone		
AUROC	0.707	0.780	0.739

Table 5. Generic model performance. Area under the ROC curve

Model	Belgium	The Netherlands	Germany
Generic - no 'country' dummy variable		0.745	
Generic - 'country' dummy variable included		0.746	
Generic - no 'country' dummy variable	0.701	0.777	0.731
Generic - 'country' dummy variable included	0.705	0.778	0.729
National	0.707	0.780	0.739
Full-data	0.722	0.800	0.789

Table 6. Number of cases predicted to be bad for cut-offs based on misclassifications of goods

	Number of cases in hold-out samples				Predicted bads by generic model 10% of total goods predicted as bad				Predicted bads by national models							
									1a) cut-offs: 10% of actual goods predicted bad by each national model				1b) cut-offs: number of goods predicted bad by each national model (cols 13-16, row 1) equals that number predicted by generic model with 10% error rate (cols 5-8, row 1)			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Bel-gium	The Ne-therlands	Ger-many	Total	Bel-gium	The Ne-therlands	Ger-many	Total	Bel-gium	The Ne-therlands	Ger-many	Total	Bel-gium	The Ne-therlands	Ger-many	Total
1. Actual goods	6935	21109	19981	48025	**481**	**2665**	**1657**	**4803**	694	2111	1998	4803	**481**	**2665**	**1657**	**4803**
2. Actual bads	923	3365	2593	6881	196	1563	735	2494	118	1383	882	2383	75	1575	772	2422
3. Total	7858	24474	22574	54906	677	4228	2392	7297	812	3494	2880	7186	556	4240	2429	7225
Difference in numbers of correctly classified bads between national models and generic model									-78	-180	147	-111	-121	12	37	-72

Table 7. Cases predicted bad for cut-off based on fixed (0.8826) probability and bad rate

	Number of cases in hold-out samples	2 a) Bads predicted by generic model	2b) Bads predicted by generic model	2a) Bads predicted by national models				Difference between national (column 7) and generic models (column 2)	2b) Bads predicted by national models				Difference between national (column 12) and generic models (column 3)
				Bel-gium	The Ne-therlands	Ger-many	Total		Bel-gium	The Ne-therlands	Ger-many	Total	
	1	2	3	4	5	6	7	8	9	10	11	12	13
Actual goods	48025	16974	4493	2461	6746	6365	15572	-1402	667	2017	1778	4462	-31
Actual bads	**6881**	4988	2388	615	2499	1742	4856	-132	256	1348	815	2419	-31
Total	54906	21962	**6881**	3076	9245	8107	20428	-1534	923	3365	2593	**6881**	-62

Table 8. Cases predicted bad for cut-offs based on marginal rates

Marginal ratio 7:1	Hold-out sample	Predicted bads by generic model	Predicted bads by national models				Difference between national (column 6) and generic models (column 2)	Predicted bads by generic model	Difference between national (column 8) and generic models (column 2)
			Belgium	The Netherlands	Germany	Total			
	1	2	3	4	5	6	7	8	9
Cut-off probabilities		0.875	0.87	0.88	0.88			Same number of misclassified as the total for national models	
Errors									
Actual goods	48025	15624	2162	6559	6221	**14942**	-682	**14942**	
Actual bads	6881	4813	573	2462	1722	4757	-56	4701	+56
Total	54906	20437	2735	9021	7943	19699	-738	19643	+56
Marginal ratio 9:1	Hold-out sample	Predicted bads by generic model	Predicted bads by national models				Difference between national (column 6) and generic models (column 2)	Predicted bads by generic model	Difference between national (column 8) and generic models (column 2)
			Belgium	The Netherlands	Germany	Total			
Cut-off probabilities		0.895	0.895	0.9	0.905				
Errors								Same number of misclassified as the total for national models	
Actual goods	48025	19517	2807	8168	8116	**19091**	-426	**19091**	
Actual bads	6881	5298	661	2684	1977	5322	+24	5251	+71
Total	54906	24815	3468	10852	10093	24413	-402	24342	+71

Table 9. Differences between the applicants accepted by one model but rejected by another one . Belgium.

Characteristic	Attribute	Total	Accepted by generic, rejected by country (3.83%)	Accepted by country, rejected by generic (3.83%)
		% cases		
Business type	Industry	18.56%	28.24%	15.61%
	Service prof	7.54%	2.99%	14.29%
Card Insurance	No Insurance	82.77%	85.71%	73.09%
	Insurance	17.18%	14.29%	26.91%
Marital Status	Single	27.09%	45.18%	64.12%
	Divorced	12.54%	16.61%	6.98%
Occupation	Part-Time	5.10%	8.31%	2.66%
	Self-Employed	3.96%	6.64%	11.96%
Telephone	Given	89.01%	92.03%	73.09%
Residential Status	Rented Flat	18.70%	28.90%	16.61%
	Living with parents	12.94%	26.25%	35.55%

	Mean			**Median**		
	Total	**Accepted by generic, rejected by country**	**Accepted by country, rejected by generic**	**Total**	**Accepted by generic, rejected by country**	**Accepted by country, rejected by generic**
Age	38	31	28	37	29	25
Time at address	7yr 9m	2yr 2m	4yr 5m	7yr 1m	2yr 10m	6yr 8m
Time on job	7yr 9m	2yr 3m	2yr 7 m	5yr 8m	1yr 6m	1yr 8m